The Study of Plato's Texts by the Use of Computers

Mari NAGASE

(1)

In the past ten years, stylometry has developed in revolutionary fashion as a result of the craving use of computers for statistical analysis. The older stylometry, developed a hundred years ago by Lewis Campbell and Lutoslawski[1], produced profitable results, many of which are being proved valid in the light of the new method. I would like to call this new way of investigation "New Stylometry", following Professor A.Q. Morton[2]. It may be defined as a study using numerical methods for the solution of literary problems. Affected by recent developments in computer studies, the British Academy has set up a central computing library for the classics[3]. One of their outstanding successes has been making a modern concordance programe which makes a statistical approach possible in various ways. Already put into practice, it is much more than the traditional word-list, or concordance. These concordances are so comprehensive as to list every word in the text, and they are arranged not only to supply the context of each word, but also to have the entries ordered by the context through a simple alphabetical list of all the words in the text. This is a useful supplement to the many indices needed for inflected languages like Greek. It also many makes possible indices arranged by word endings, i.e. a reverse index in which the words are placed by the last letters, not their first letters[4]. Moreover it is a simple matter to record not only the word form, but also the groups or patterns of words, of stress, of combinations of vowels or consonants, of grammatical structures, or linguistic elements[5]. It is also

possible to record and analyze the occurrence of the different sizes or kinds of those events, for example, how many sentences have ten words in them, or how many have four clauses. Further, it is simple to record the spacing between the occurrences of events.

The computer is, in principle, a simple tool; it is rather like a library or office filing system in that it has two parts, a storage section where information is held, and an operational section which counts, sorts, and arranges the stored information. The great advantages of the computer lie in the small storage size needed, and its speed of operating on the stored information. Also, it can be arranged very simply, so that material relevant to the testing of any hypothesis can be brought out and put before the scholar with minimum effort. Nowadays its use covers the whole range of scholarly activities. It can be used for literary, historical, and philosophical investigations. It can examine the language of an epoch, school, or an individual; it can be used for linguistic analysis, for solving problems of authenticity of integrity and chronology, and for doing research aimed at uncovering knowledge or in teaching how to simplify old things. Literally speaking, there can be as many uses as users[6].

(2)

In the field of Greek Studies, the surveys were, in the early stages, devoted to vocabulary, especially uncommon words or once-occurring words. This study was based on the axiomatic assumption that every author has a store of material in his mind and that once-occurring words are those newly emerged from this store. If such a store of words is analysed, it will tell how words are stored in and then issue them from the brain or mind. Later scholars have, however, concentrated on the study of small and common words, like particles and connectives, which supply us with a large quantity of data. The study of high-frequency words like this is quite effective when there is no external evidence and

a decision must be made entirely on the internal evidence supplied by texts. In addition, these common words are not affected by the subject of the text[7]. That is to say, these words may be assumed to reflect the unconscious habits of the writer more clearly than uncommon words.

It should be declared here that an investigation must be based on the counting of repeated events and that this counting must be analysed mathematically. This is the foundation of any such research. It is worth-while, also to observe De Morgan's suggestion, for he established the foundation of modern stylometry[8]. He set out three basic principles for a scientific approach. In the first place, the approach should be general and objective. In the second place, the statement about language must be statistical. It would for example, be nonsense to say that in one thousand words Plato uses μέν 50 times; it might be possible to say only that, on average, Plato uses μέν 50 times in one thousand words of a certain text. All descriptions of habits and comparisons of habits must be averages or other statistics which show the same characteristics as averages. Thirdly a judgement about a text should be made on the basis of the whole text, or as much of it as can be included in the tests, not on a few isolated sections which may or may not represent the whole.

Most of the analysis is carried out with the concept of the mean and variance, which are very common. Also, there are several ways of analysis invented recently, like multivaliated analysis, cluster analysis[9], factor analysis, multidimensional scaling, and semantic-distance method. Basically these methods are derived from the idea of quality could be interpreted to quantity.

(3)

After the brief outline of the development of stylometry, it would be better for us to proceed to the new project which is now under way.

At present, scholars have started to pay more attention to experiments dealing with aspects of sentence structure rather than with vocabulary. In the above description, I use the word a "project", because it is essential to have a team consisting of a statistical analyst, a computer programmer, and a classical scholar when any persues research with computers. To be fully qualified to undertake this kind of research alone, one must be a professional philosopher, classicist, and statistician at the same time.

From October 1st, 1980, to March 31st, 1982, I joined the project as a visiting researcher of the Department of Greek and Latin in the University of Manchester, sponsored by the British Council. This project to study Plato's prose style was organised in 1979 by Dr. L. Brandwood and Mr. G. Neal from the University of Manchester, Mrs. P. Hubby from the University of Liverpool, and Mrs. B. Mandl from the Regional Computer Center of the University of Manchester.

The puopose of this project are divided into three items; first, to investigate some specific features of Plato's prose style, aiming to produce a certain, quantitative measure of sentence structure, which involves the establishment of a method or principles for comparing the texts written by Plato; second, to apply the method to draw some profitable results from Plato's prose style and third, to offer a grammatical background to Plato's sentence structure and to show habitual usages which are peculiar to Plato, so that one can follow his writings easily and objectively. This last aim is connected with the investigation of the automatic analysis system, which is currently being attempted in other languages as well.

Concurrently our method is inductive. We had to repeat several experiments in order to establish effective rules for distinguishing the differences in author's styles. At the beginning it is requisite to make some axiomatic assumptions.

As we have seen in the introductory explanation, many features has been proposed to be significant for discriminating differences in the styles

of different texts. Scholarly attention has, however, been paid mainly to vocabulary studies[10]. No attempt has yet been made to examine the complexities of sentence structures.

To establish new principles of discrimination, we ourselves, have therefore, concentrated on sentence studies. The main reason why the study of the complexities of sentences is promising is that they are free from fluctuations due to change in subject matter. On the other hand, vocabulary is sometimes chosen according to the subject matter. Once-occurring words and uncommon words especially are easily influenced from the start by the subject. Accordingly, such words do not supply us with sufficient data reflecting unconscious personal stylistic features.

At this point, our basic assumption raized the new question that there might be personal peculiarities in the level of sentence construction. To check sentence construction and the complexities of sentences, there are many possible ways. After several experiments, we selected some pseudo-criterion, according to which new programs were written, originally by Mr. G. Neal.

The lists of points to check may be summarized as follows:

1. Count clauses by kind and species. There are four kinds: ⓐ main and parenthetical clause, ⓑ subordinate clause, ⓒ infinitive phrase, and ⓓ participial phrase. The specific divisions are also four: ⓐ simple, ⓑ coordinate, ⓒ split, and ⓓ split/coordinate.
2. Count clauses by the level of subordination.
3. Count clauses by complexity (i.e., by number of ⓐ subordinate clause and ⓑ other syntagms in clauses
4. Count syntagems by (simplified) type and species. (Lists of syntax codes will be given later, with a definition of the coding scheme.)
5. Keep track of the maximum depth of subordination encountered in a given sentence.

6. Count number of subordinate clauses and of ordinary syntagms in the current clause.
7. Count total numbers of syntagms and words in the data.
8. Record the last suntagms met with in a clause.
9. Record, separately, the number of modules met with in any sentence in the data so far. Variables are used in the analysis of the specific distinctions between different varieties of subordinate clauses and other syntagms.

Obviously we made these lists in order to check the syntagms and the position of syntagms which relate to sentence construction deeply. A more precise explanation will follow, togather with the explanation of each table. To facilitate making programs, one program combines two or more checking items.

(4)

Before listing the tables, however, it might be convenient to give some explanation of our magnetic tape and of our coding and coding systems. To get overall view of the operation as it is at present constituted may be useful as well when the reader uses computers for his own research, although some of the following explanation will be limited to the very specific range of our work.

1. Data

Because the Greek text of Plato was already available to us on magnetic tape (from the previous work which Dr. Leonard Brandwood did in compiling "A Word-Index to Plato"), and because we are able to obtain a tape of Xenophon from the Oxford Archive of the Oxford University Computer Service, we decided to feed our syntactic coding into the computer from the text which it relates to. Among Plato's writings, we chose three texts, "Lysis", "Theaetetus" and "Philebus" which are supposed to be written in three different period, early, middle,

and later time of his life. From Xenophon's text, "Memorabilia" was chosen.

Each text is divided into four sections according to the computer programs as follows:

(1) "Laches"; (Plato)

① Section 1 178A1—184C8
② Section 2 184C —190A8
③ Section 3 190A —196A3
④ Section 4 196A —201C2

(2) "Theaetetus"; (Plato)

① Section 1 142A1—145C7
② Section 2 145C —148E6
③ Section 3 148E —154D7
④ Section 4 154D —158E11

(3) "Philebus"; (Plato)

① Section 1 11A1—17A5
② Section 2 17A —22B9
③ Section 3 22B —27B9
④ Section 4 27B —33E13

(4) "Memorabilia"; (Xenophon)

① Section 1 2. 1. 1. 1—2. 2. 1. 4
② Section 2 2. 2. 1 —2. 2. 12. 5
③ Section 3 2. 2. 12 —2. 7. 1. 3
④ Section 4 2. 7. 1 —2. 10. 6. 4

2. Definition of coding scheme[11]

The coding system described below is intended for input to the computer *separately* from the text to which it applies. It therefore uses the symbol > as a word-marker. Text and coding may then be easily married by programming the computer to insert one word from the text at each > in the coding.

The sentence is analysed for the purposes of this scheme into clauses (here regarded as including infinitive and participial phrases as well as main and subordinate clauses with finite verbs). It is important to note that the method of analysis used treats dependent clauses as *part of* their governing clause; each main clause is, therefore, analysed as a basic independent unit of the text, but dependent clauses are regarded as nested, within each other.

The same basic code-pattern is used for a clause at any level of subordination, but the different types of clause are indicated by the clause code with which the pattern begins.

CLAUSE-PATTERN

(a) Clause Code (see list below), but missing in the case of second or subsequent parts of a clause that is split by the fact that members of governing clauses intervene.
(b) Space.
(c) A succession of Syntax Codes, until the end of the clause (or its current part) is reached. See below for details.

[LIST OF CLAUSE CODES]

Main

101 Statement
102 Question
103 Command
104 Wish
105 Exclamation

Subordinate

10 Substantival
11 Relative clause with antecedent omitted
15 Adjectival
20 Result

22 Place
23 Cause
24 Manner
25 Time
26 Purpose
27 Conditional
28 Means
29 Concession

Infinitive

30 Without article
31 With article

Participial

40 Miscellaneous (incl. indirect statement)
41 With article (=substantive)
42 Adjectival
43 Cause
44 Manner, means
45 Time
46 Purpose
47 Condition
48 'Narrative' or other participle logically coordinated with main verb
49 Concession

[LIST OF SYNTAX CODES]

0	Subject
1	Verb
2	Direct object or equivalent
3	Indirect object or equivalent
4	Predicate
5, 6, 7, 8, 9	Attribute (respectively) of 0, 1, 2, 3, 4
55, 555, 5555, etc.	Attribute of unit with code consisting of

	one less 5
66, 777, 8888, 99999, etc.	Attribute of unit with code consisting of one less digit of the same value. (66 may also indicate an attribute of any of the 60—69 codes listed below)
60	Simple negative (if marked)
61	Compound negative (if not coded according to a more specific function)
62	Adverb
63	Prepositional phrase
64	Oath formula
67	Not used in coding at the moment, but when it appears in processing it indicates words at *start* of clause not specifically coded.
68	Conjunction, particle (if present)
69	Vocative (if not coded as attribute of some other unit)

[LIST OF PHRASE CODES]

0(	Subject clauses or phrases
2(	Direct-object clauses or phrases
3(	Indirect-object clauses or phrases
4(	Predicate clauses or phrases
5(, 7(, 8(, 9(	Attribute clauses or phrases (respectively) of 0, 2, 3, 4
62(	Adverbial clauses or phrases
63(	Prepositional clauses or phrases

3. Merging

Picking up the appropriate numbers from this coding scheme, we tagged them to word, phrase and clause in the four texts[12]. Then the

coding was converted from the original form in which it was prepared and fed into the computer to a modular version. In this version two things were arranged: ① clauses were separated from the governing clauses in which they were emended, so that they can be treated as separate units in subsequent processing, ② various linked items were tagged with the module number in which the details of the related items are recorded.

4. Result

An investigation will be based on the counting of repeated words, phrases and clauses, and counting will be made in suitable units[13]. Among many criterions, the means and variance are the most important descriptive charactors of set of our observations. Also statistics are produced in the form of a frequency distribution and a matrix which contains figures concerning the relationship between the data. Some supplementary ratios are calculated as well.

(5)

Turning from the procedural explanation, our main inquiry should now be undertaken.

The first step is to discover the number of occurrences of clauses by kind and species.

Table 1 sets out the results in the forms of both of numbers and percentages. The four sections in each text are coordinated to the four sections described in the Explanation of Data. The number in parentheses in the section of the participial phrases shows the number of participles used especially as Adverbial Participial Phrases; they are tagged by the coding numbers of 43 (Cause). 44 (Manner, means), 45 (Time), 46 (Purpose), 47 (Condition), and 49 (Concession). These

Table 1

Clause Totals

"Laches"

Type / Section	Main Clause	%	Adverbial Subordinate Clause	%	Infinitive Phrase	%	Participial Phrases	%	Total Clause
①	126	29.1	62	14.3	98	22.6	74 (47)	17.1 (10.8)	433 (77)
②	130	28.3	53	11.5	102	22.2	70 (33)	15.2 (7.2)	460 (214)
③	215	43.7	41	8.3	98	19.9	57 (30)	11.6 (6.0)	492 (105)
④	162	37.9	44	10.3	108	25.2	57 (17)	13.3 (4.0)	428 (93)

"Theaetetus"

Type / Section	Main Clause	%	Adverbial Subordinate Clause	%	Infinitive Phrase	%	Participial Phrases	%	Total Clause
①	122	45.5	29	10.8	41	15.3	49 (31)	18.3 (11.6)	268 (65)
②	114	44.4	19	7.4	44	17.1	35 (21)	13.6 (18.2)	257 (60)
③	195	42.1	33	7.1	73	15.8	78 (42)	16.8 (9.1)	463 (92)
④	105	30.9	35	10.3	64	18.8	71 (25)	20.9) (7.3)	340 (127)

"Philebus"

Type / Section	Mein Clause	%	Adverbial Subordinate Clause	%	Infinitive Phrase	%	Participial Phrases	%	Total Clause
①	154	35.0	39	8.9	104	23.6	92 (51)	20.9 (11.6)	440 (141)
②	146	38.7	28	7.4	71	18.8	80 (44)	21.2 (11.7)	377 (143)
③	176	48.2	22	6.0	50	13.7	75 (36)	20.5 (9.9)	365 (101)
④	215	45.6	36	7.6	70	14.9	92 (41)	19.5 (8.7)	471 (117)

"Memorabilia"

Type / Section	Nain Clause		Adverbial Subordinate Clause		Infinitive Phrase		Participial Phrases		Total Clause
		%		%		%		%	
①	172	28.2	62	10.2	143	23.5	151 (69)	24.8 (11.3)	609 (213)
②	52	24.9	23	11.0	49	23.4	46 (30)	22.0 (14.3)	209 (153)
③	257	24.9	108	10.4	253	24.5	234 (139)	22.6 (13.4)	1,034 (144)
④	141	28.3	49	9.8	112	22.5	93 (57)	18.7 (11.4)	498 (106)

numbers are important in calculating correlation of subordinate clauses and the other phrases.

A study of Table 1 reveals that the frequency of adverbial subordinate clauses decreases in the order of "Laches", "Theaetetus" and "Philebos". In Xenophon's "Memorabilia", it is evident that the usage of main clausesis is very low compared to Plato's dialogues. Also, we must take a further look to see if the proportions of infinitive phrases and participial phrases are different among Plato's texts. In "Laches" it is remarkable that infinitive phrases are used more often than participial phrases. In "Theaetetus", the difference between these rates is small, but in "Philebus", except in Section 1, the result is the opposite of that of "Laches". In this text, participial phrases are used more often than infinitive phrases.

From this results, we could presume that there might be a close relation between the rate of the usage of adverbial subordinate clauses and participial or infinitive phrases.

If we look back to the historical development of the Greek Language it is commonly accepted that in the 5th century B.C., participles are scarcely used. But the language has developed rapidly within 150 years. According to the whole-scale survey of Greek prose writings checked by the Oxford Computer Centre, it is reported that the usage of participle

increases around 440 B.C.[14]. This would suggest also that the complexity of subordinate clauses is related to the development of the usage of participles. Using the value of adverbial aubordinate participial phrases, I calculated the Coefficient of Correlation, which measures how much two series of observations are connected. I checked the correlation between the subordinate clauses and the adverbial subordinate phrases. As well, I counted the correlation of subordinate clauses and infinitive phrases for the sake of comparison.
The results are as follows:

Table 2
Coefficient of Correlation

CORRELATION / Texts	Subordinate Clause & Adverbial Subordinate Participle	Subordinate Clause & Infinitive Phrases
Laches	0.8	0.3
Theaetetus	−0.4	−0.9
Philebos	0.4	0.9
Memorabilia	0.9	0.4

Correlation (r) takes a value from -1 to $+1$:

$$-1 \leqq r \leqq +1$$

The nearer the value of (r) is to $+1$ or -1, the closer is the relation of the two series of numbers. On the contrary, if the value of (r) is near to 0, this means that their relation is less. $r > 0$ means the relationship is which one increases and the other increases also. $r < 0$ means the relationship is which one increases while the other decreases.

An inspection of Table 2 suggests that the values of the correlation of subordinate clauses and adverbial participles are high in the cases of "Laches" and "Memorabilia".

Judging from this point, the usage of adverbial subordinate participial phrases can be said to be similar in both texts. However, we cannot

yet be said to have satisfactorily guantified the similarity of the two texts, for the statistical procedures which have been adopted are strictily applicable in only one aspect. We need more experiments.

Table 3 gives the rate of frequency of sentences by maximum depth of subordination. The test was intended to determine the complexity of sentences.

The first column shows various patterns of sentences which differ from each other according to the number of nested clauses and phrases.

Table 3

Frequency of Sentences by Maximum Depth of Subordination

Conventions Used Here: CL=Maximum Number of Nested Clauses in Sentence
PH=Maximum Number of Nested Infinitive or Participial Phrases
Results=Number of Occurences of Various Sentence Patterns and Percentage

"Laches"

Section ①

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	16	15.7	1+0	15	14.7	2+0	5	4.9	—	—	—
0+1	18	17.6	1+1	17	16.7	2+1	7	6.9	3+1	2	2.0
0+2	5	4.9	1+2	10	9.8	2+2	2	2.0	3+2	1	1.0
0+3	1	1.0	1+3	1	1.0	2+3	1	1.0	—	—	—
									3+4	1	1.0

Section ②

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	50	49.0	1+0	14	13.7	—	—	—	3+0	3	2.9
0+1	10	9.8	1+1	11	10.8	—	—	—			
0+2	3	2.9	1+2	8	7.8	2+2	1	1.0			
0+3	2	2.0									

Section ③

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	94	48.5	1+0	22	11.3	2+0	2	1.0	3+0	3	1.5
0+1	23	11.9	1+1	16	8.2	2+1	6	3.1	3+1	1	0.5
0+2	8	4.1	1+2	12	6.2	2+2	1	0.5			
			1+3	2	1.0	2+3	3	1.5			
			1+4	1	0.5						

Section ④

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	55	39.0	1+0	23	16.3	2+0	2	1.4
0+1	16	11.3	1+1	17	12.1	2+1	3	2.1
0+2	5	3.5	1+2	12	8.5	2+2	4	2.8
0+3	1	0.7	1+3	1	0.7	—	—	—
			1+4	1	0.7	2+4	1	0.7

"Theaetetus"

Section ①

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	40	41.7	1+0	14	14.6	2+0	3	3.1	3+0	1	1.0
0+1	18	18.7	1+1	11	11.5	2+1	1	1.0			
0+2	2	2.1	1+2	3	3.1						
0+3	2	2.1	1+3	1	1.0						

Section ②

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	50	49.0	1+0	14	13.7	—	—	—	3+0	3	2.9
0+1	10	9.8	1+1	11	10.8	—	—	—			
0+2	3	2.9	1+2	8	7.8	2+2	1	1.0			
0+3	2	2.0									

Section ③

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	76	47.5	1+0	18	11.2	2+0	6	3.7				4+0	1	0.6
0+1	16	10.0	1+1	25	15.6	2+1	3	1.9						
0+2	6	3.7	1+2	7	4.4	—	—	—						
			1+3	1	0.6	2+3	1	0.6						

Section ④

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	27	31.0	1+0	13	14.9	2+0	3	3.4	3+0	2	2.3	4+0	1	1.1
0+1	8	9.2	1+1	18	20.7	2+1	5	5.7						
0+2	2	2.3	1+2	4	4.6	—	—	—						
			1+3	3	3.4	2+3	1	1.1						

"Philebus"

Section ①

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	44	33.6	1+0	15	11.5	2+0	7	5.3	3+0	1	0.8	—	—	—
0+1	28	21.4	1+1	13	9.9	—	—	—	3+1	1	0.8	4+1	1	0.8
0+2	6	4.6	1+2	5	3.8	2+2	3	2.3	—	—	—			
0+3	4	3.1	1+3	2	1.5				3+3	1	0.8			

Section ②

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	53	41.4	1+0	14	10.9	2+0	3	2.3	—	—	—
0+1	23	18.0	1+1	15	11.7	2+1	1	0.8	3+1	2	1.6
0+2	7	5.5	1+2	5	3.9	2+2	1	0.8			
—	—	—	1+3	2	1.6						
0+4	1	0.8									

Section ③

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	64	44.4	1+0	20	13.9	2+0	2	1.4	3+0	1	0.7
0+1	30	20.8	1+1	13	9.0	—	—	—	3+1	1	0.7
0+2	5	3.5	1+2	3	2.1	2+2	1	0.7			
0+3	2	1.4	1+3	1	0.7	2+3	1	0.7			

Section ④

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	95	48.7	1+0	23	11.8	2+0	1	0.5	—	—	—
0+1	36	18.5	1+1	16	8.2	2+1	4	2.1	3+1	1	0.5
0+2	4	2.1	1+2	10	5.1				3+2	1	0.5
0+3	4	2.1									

"Memorabilia"

Section ①

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	26	19.7	1+0	23	17.4	2+0	3	2.3	3+0	1	0.8	—	—	—
0+1	16	12.1	1+1	20	15.2	2+1	6	4.5	3+1	1	0.8	—	—	—
0+2	6	4.5	1+2	7	5.3	2+2	4	3.0	3+2	2	1.5	4+2	1	0.8
0+3	7	5.3	1+3	4	3.0									
0+4	2	1.5	1+4	3	2.3									

Section ②

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	5	11.4	1+0	8	18.2	2+0	1	2.3	—	—	—
0+1	4	9.1	1+1	11	25.0	2+1	3	6.8	—	—	—
—	—	—	1+2	6	13.6	2+2	3	6.8	3+2	1	2.3
0+3	1	2.3				2+3	1	2.3			

Section ③

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	29	12.8	1+0	43	19.4	2+0	6	2.7	3+0	2	0.9	—	—	—
0+1	30	13.3	1+1	23	10.2	2+1	12	5.3	3+1	4	1.8	4+1	1	0.4
0+2	15	6.6	1+2	19	8.4	2+2	19	8.4	3+2	3	1.3			
0+3	9	4.0	1+3	4	1.8	2+3	4	1.8						
			1+4	3	1.3									

Section ④

CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%	CL+PH	No.	%
0+0	15	12.6	1+0	33	27.7	2+0	5	4.2	3+0	2	1.7	—	—	—
0+1	11	9.2	1+1	18	15.1	2+1	10	8.4	—	—	—	—	—	—
0+2	7	5.9	1+2	3	2.5	2+2	6	5.0	3+2	1	0.8	4+2	1	0.1
0+3	1	0.8	1+3	5	4.2									
—	—	—												
0+5	1	0.8												

The second column gives cumulative totals of occurences in each section. The third column shows the percentage, so that one can see at a glance how many sentences are accounted by forms of different clauses and phrases. The variety of sentence pafferns are different in each section.

From this table we counted the percentage of sentences consisting of more than two clauses and more than three clauses. The means of sentences which have more than two clauses are 10.9%, 7.35%, 6.0% and 19.4% in "Laches", "Theaetetus", "Philebos", and "Memorabilia" respectively. The means of sentences having more than three clauses are 2.63%, 1.96%, 1.80%, and 3.48% in "Laches", Theaetetus", "Philebos", and "Memorabilia" respectively. The first fact which meets the eye in studying these numbers is that, in Xenophon's text, the percentage of sentences which have more than two clauses is very high compared with the dialogues written by Plato. Among Plato's writings,

only "Laches" has a high frequency. It can thus be said Xenophon used complex sentences more often than Plato did. From the rate of usage of complex sentences, the position of "Laches" can be said to be intermediate between Xenophon's "Memorabilia" and Plato's two other texts.

The argument could be carried a stage further if we could check the relationship between clauses or phrases so as to show which clauses depend on the main clause or adverbial clauses, or which clause is related to the subject or adjective phrases, etc. The least we can say at this stage is that this method is effective in checking the complexity of sentences by, for instance, counting the levels of subordination for the purpose of discriminating author's styles.

Table 4 shows the totals of ordinary syntagms, while Table 5 shows the totals of syntagms consisting of clauses or phrases. The results are represented in percentages. From the inspection of Table 4 we could make some significant observations:

1. From a comparison of the frequency of verbs, the value of the usage of verbs is uniform among Plato's writings, but different from that of Xenophon.
2. The value of an attribute of subject is lower in "Memorabilia", but higher in "Philebus".
3. The value of an attribute of verb in "Philebus" is significantly different from those in other texts.
4. The frequency of prepositions is lower in "Memorabilia" than in the dialogues written by Plato.

The most striking feature of Table 5 is that the frequency of an attribute of subject, coded 5(, is lowest in the "Laches" among Plato's texts. In the case of clauses and phrases, the comparison of the attribute of subject is less satisfactory.

Taking togather the data presented in Table 4 auf Table 5, the frequency of syntagms may be seen to provide us with much information

Table 4

Totals of. Ordinary Syntagms (Excludig Clauses/Phrases)

"Laches"

Type / Section	0	1	2	3	4	5	6	62	63	7	8	9
	%	%	%	%	%	%	%	%	%	%	%	%
①	11.7	32.5	11.0	3.7	5.9	4.4	7.4	8.0	8.2	4.7	0.7	1.7
②	9	33.4	8.6	5.5	7.0	3.4	11.2	8.6	7.3	3.2	0.3	2.5
③	12.8	31.3	7.9	3.6	8.5	4.3	11.5	7.5	6.0	2.0	0.8	3.7
④	13.2	32.3	9.8	4.2	6.1	3.9	10.6	8.8	4.0	3.4	1.3	2.4

"Theaetetus"

Type / Section	0	1	2	3	4	5	6	62	63	7	8	9
	%	%	%	%	%	%	%	%	%	%	%	%
①	8.9	34.0	8.3	4.9	6.6	4.0	8.6	12.1	5.1	2.5	1.4	3.7
②	11.7	29.9	9.9	3.4	8.0	3.3	10.5	7.0	4.6	6.4	1.6	3.8
③	13.2	30.7	8.9	4.9	9.1	4.8	10.2	6.4	4.6	1.9	1.3	4.0
④	13.4	31.1	8.0	3.0	8.2	5.3	11.5	8.0	5.2	3.5	0.2	2.8

"Philebus"

Type / Section	0	1	2	3	4	5	6	62	63	7	8	9
	%	%	%	%	%	%	%	%	%	%	%	%
①	12.3	30.8	9.6	2.1	10.1	7.3	4.2	8.3	5.0	5.2	0.6	4.4
②	10.6	29.5	10.8	3.4	8.4	6.0	6.7	8.7	7.7	5.3	0.7	2.3
③	8.7	31.9	15.1	3.3	7.4	4.1	6.0	8.3	6.7	6.2	6.4	1.9
④	10.2	30.1	9.3	3.0	9.3	5.9	6.5	8.6	7.6	4.1	0.5	5.0

"Memorabilia"

Type / Section	0	1	2	3	4	5	6	62	63	7	9	9
	%	%	%	%	%	%	%	%	%	%	%	%
①	12.3	37.4	10.2	4.1	4.6	3.9	9.6	3.7	6.3	3.7	0.8	3.5
②	11.7	39.5	9.3	4.8	5.2	1.8	10.5	5.0	4.4	4.4	0.6	2.6
③	12.3	37.3	10.1	3.7	7.0	2.9	10.1	6.1	4.8	2.2	0.3	3.2
④	11.9	38.1	11.3	3.7	5.8	2.8	8.6	6.8	5.6	3.3	0.2	2.0

Table 5

Totals of Syntagms Consisting of Clauses/Phrases

"Laches"

Type / Section	0(	2(	3(	4(	5(	6(	63(	7(	8(	9(
	%	%	%	%	%	%	%	%	%	%
①	17.9	26.0	0	2.3	16.0	28.2	0.4	5.3	0.4	3.4
②	13.2	32.4	0	5.9	11.1	24.2	0	6.6	2.1	4.2
③	12.6	36.2	0.8	6.3	14.2	19.3	1.6	3.1	1.6	4.3
④	13.9	35.7	0.9	7.0	7.0	24.8	0.4	5.7	1.7	3.0

"Theaetetus"

Type / Section	0(	2(	3(	4(	5(	6(	63(	7(	8(	9(
	%	%	%	%	%	%	%	%	%	%
①	7.9	30.7	0.7	4.3	20.7	22.9	0	4.3	4.3	4.3
②	12.3	33.8	2.3	7.7	14.6	17.7	0	7.7	0.8	3.1
③	14.2	32.9	0.9	7.1	17.8	20.4	0	3.6	0.4	2.7
④	17.5	29.0	2.5	9.5	14.5	19.5	0.5	4.5	1.0	1.5

"Philebus"

Type / Section	0(	2(	3(	4(	5(	6(	63(	7(	8(	9(
	%	%	%	%	%	%	%	%	%	%
①	14.5	28.5	1.2	4.4	19.7	21.3	0.4	6.0	0.8	3.2
②	14.4	29.4	1.0	5.5	26.4	17.4	0	3.5	1.0	1.5
③	14.7	28.8	0	6.2	19.2	19.8	0.6	8.5	0.6	1.7
④	19.2	26.0	1.4	3.7	18.7	21.9	0.5	5.0	0.9	2.7

"Memorabilia"

Type / Section	0(	(2	3(	4(	5(	6(	63(	7(	8(	9(
	%	%	%	%	%	%	%	%	%	%
①	14.2	34.6	2.5	2.5	16.5	21.2	0.6	4.2	1.7	2.0
②	6.6	45.3	0.7	4.4	18.2	17.5	0	2.9	0.7	3.6
③	11.3	40.3	1.1	3.0	18.7	18.4	0.2	4.2	0.8	2.1
④	11.3	41.0	0.6	2.3	17.1	17.1	0	6.1	1.9	2.6

for our purpose of discriminating the styles of texts.

In respect of the frequency of syntagms of verbs, attribute of subject, and prepositions, we can discriminate Xenophon's texts from Plato's writings. Further, except from the irregular rate of occurrence in Section 4 (7.4%), "Laches" is somewhat similar to "Memorabilia". We have reached the same result in the first test. Here again, we could obtain a strong prediction of similarity between "Laches" and "Memorabilia".

Table 6

Totals of Coordination & Split %

"Laches"

Type / Section	0 : Subject		0(: Subject Clause		2 : Direct Object		2(: Direct Object Clause		63 : Prepositional Phrase	
	Coordinate	Split	Coordinate	Split	Coordinate	Split	Coordinate	Split	Coordinate	Split
	%	%	%	%	%	%	%	%	%	%
①	9.5	0.7	0	10.6	6.5	1.4	0	25.0	6.7	13.5
②	7.0	0.9	0	28.9	8.3	3.7	3.2	22.6	9.7	5.4
③	3.2	6.3	0	29.0	7.1	5.1	5.4	20.7	27.0	18.9
④	16.6	3.8	0	21.9	11.1	1.7	1.2	29.3	4.7	0

"Theaetetus"

Type / Section	0 : Subject		0(: Subject Clause		2 : Direct Object		2c : Direct Object Clause		63 : Prepositional Phrase	
	Coordinate	Split	Coordinate	Split	Coordinate	Split	Coordinate	Split	Coordinate	Split
	%	%	%	%	%	%	%	%	%	%
①	3.4	8.6	18.2	9.1	7.4	0	0	16.3	27.3	6.1
②	8.9	0	18.7	6.2	6.0	6.0	0	27.3	32.3	6.5
③	20.6	3.6	12.5	12.5	19.8	4.5	5.4	20.3	29.8	0.0
④	19.1	3.8	14.3	2.9	11.5	3.8	5.2	24.1	25.5	7.8

"Philebus"

Section \ Type	0 : Subject		0(: Subject Clause		2 : Direct Object		2(: Direct Object Clause		63 : Prepositional Phrase	
	Coordinate	Split	Coordinate	Split	Coordinate	Split	Coordinate	Split	Coordinate	Split
	%	%	%	%	%	%	%	%	%	%
①	11.8	8.3	25.0	19.4	11.5	3.5	11.3	23.9	10.2	11.9
②	15.5	5.5	0	27.6	21.4	0.9	8.5	18.6	13.7	12.5
③	19.0	4.8	26.9	23.1	27.6	4.8	3.9	33.3	20.3	21.9
④	22.0	10.2	9.5	14.3	16.5	6.1	1.8	36.8	7.4	17.9

"Memorabilia"

Section \ Type	0 : Subject		0(: Subject Clause		2 : Direct Object		2(: Direct Object Clause		63 : Prepositional Phrase	
	Coordinate	Split	Coordinate	Split	Coordinate	Split	Coordinate	Split	Coordinate	Split
	%	%	%	%	%	%	%	%	%	%
①	6.8	4.2	3.9	19.6	5.1	7.6	0	28.2	4.1	10.3
②	0	0	0	11.1	4.3	0	0	37.1	0	13.6
③	5.1	1.9	4.0	18.7	7.4	3.1	3.0	40.1	4.0	11.3
④	10.5	4.9	5.7	31.4	14.1	3.7	0.8	38.6	3.0	7.5

Table 6 shows the totals of the coordination and split in syntagms and clause or phrase syntagms. The coordination means the case that words, clauses or phrases are used in a correlated way. The split means that there are splits between two related words. Coordination and splits are techniques of rhetoric which may differ among writings.

It leaps to the eye that, in "Philebus", that the totals of the coordination of subject are considerably high compares with the values of the other texts. Proceeding to the subject clause, the frequency of coordination is low, but the frequency of split is very high, in both "Laches" and "Memorabilia". It may be noted that, in "Memorabilia", the coordination of prepositions occurs very much less than in any other of Plato's books. Here again, there is some evidence that "Memorabilia" resembles Plato's "Laches".

Table 7 compares the frequency of syntagms appearing *first* in a clause or sentence. If one counts the four highest frequencies of such syntagms in Plato's dialogues, the order is this: the first syntagm is subject, the second is verb, the third is direct object and the fourth is adverb.

In Xenophon's "Memorabilia", however, the fourth is not adverb, but attribute of verb. A close examination, then, of the section of "Memorabilia" shows that sentences begin with object or object clauses more often than in Plato's writings.

(6)

Meditating on these results, it seems that some conclusions can be drawn.

Checking the fsequency of participles and infinitives, which clearly reflect differences in style is effective in the study of problems of both authenticity and chronology. From the grammatical point of view, there might be a close relationship between subordinate clauses and the usage of participles. We may obtain more profitable results if we apply this test to other Greek prose writers. At any rate, the Coefficient of Correlation is quite helpful for discriminating between texts.

We may assume that, it would be a useful criterion for showing the difference of style to check texts in terms of depth of complexity. However, a still more sophisticated method is needed before we can reach such discrimination as presented above. If we could determine not only the level of subordination, but also the relationaship of subordination among clauses, we could possibly achieve some information about the mechanism of sentence construction.

The most informative of the many syntagms, the frequency of the verb, the attribute of subject, and prepositions, are effective for discriminatory work. In the majority of our tests the results show the difference between Plato's writings and Xenophon's writing clearly. It is interesting to note, though, that in some tests there is more similarity between

Table 7

Frequency of Syntagms Appearing *First* in a Clause or Sentence

"Laches"

Type of Syntagms / Section	0	0(	1	2	2(	3	3(	4	4(	5
	%	%	%	%	%	%	%	%	%	%
①	13.8	2.5	19.0	14.1	4.4	3.5	0	6.7	0.7	2.5
②	12.8	2.9	16.7	9.4	5.3	4.6	0	0.5	0.5	1.9
③	17.1	2.4	12.0	10.7	6.3	4.1	0.5	1.5	1.5	1.7
④	20.6	1.6	13.6	11.0	5.6	3.7	0	0.5	0.5	0.8

"Theaetetus"

Type of Syntagms / Section	0	0(	1	2	2(	3	3(	4	4(	5
	%	%	%	%	%	%	%	%	%	%
①	10.9	1.3	17.4	6.1	3.5	5.7	0	7.8	0.4	3.0
②	14.8	1.7	16.1	9.6	5.7	3.5	0	9.6	0.9	2.6
③	19.5	2.3	13.7	11.4	6.6	4.1	0.5	7.4	0.3	1.3
④	22.0	5.1	13.2	11.1	4.4	3.0	0.3	4.7	0.3	1.7

"Philebus"

Type of Syntagms / Section	0	0(	1	2	2(	3	3(	4	4(	5
	%	%	%	%	%	%	%	%	%	%
①	17.7	3.7	17.5	13.5	4.8	2.0	0	8.2	0.8	3.4
②	13.5	3.0	14.9	16.5	4.3	1.3	0	7.3	2.0	2.6
③	11.5	3.8	14.6	15.3	6.9	2.8	0	8.0	1.0	0.3
④	12.0	4.3	14.7	12.0	6.8	2.4	0.3	7.6	0.8	3.0

"Memorabilia"

Type of Syntagms / Section	0	0(	1	2	2(	3	3(	4	4(	5
	%	%	%	%	%	%	%	%	%	%
①	21.8	4.4	15.6	11.1	9.3	2.9	0	2.9	0.5	2.2
②	14.0	2.2	14.5	9.7	15.1	3.2	0	4.8	0	2.2
③	18.4	3.0	12.7	11.3	13.0	2.3	0.2	0.2	4.9	0.4
④	17.1	2.2	16.9	8.9	12.4	2.7	0	5.8	0.2	0.4

5)	6	6(	62	63	63(	7	7(	8	8(	9	9(	67
%	%	%	%	%	%	%	%	%	%	%	%	%
2.5	4.4	5.4	10.4	6.7	0	2.0	0.5	0.5	0	0.7	0.2	0
1.4	9.9	6.8	11.8	6.8	0	1.2	0.2	0	0.5	1.7	0	0
2.2	12.7	1.7	9.3	3.9	0.2	0.5	0.2	0.5	0.2	3.2	0.7	0
0.8	13.9	4.0	9.4	3.5	0	1.6	0.8	0.3	0	0.8	0.5	0

5)	6	6(	62	63	63(	7	7(	8	8(	9	9(	67
%	%	%	%	%	%	%	%	%	%	%	%	%
6.5	9.6	3.9	15.2	3.0	0	1.7	0.9	0.9	0	2.2	0	0
2.6	13.9	2.6	7.8	3.9	0	1.7	0.4	0	0.4	2.2	0	0
2.3	10.2	4.1	7.4	4.6	0	1.8	0	0.3	0	1.5	1.0	0
1.4	13.2	2.7	7.4	5.4	0	2.7	0.3	0.3	0	0.7	0	0

5)	6	6(	62	63	63(	7	7(	8	8(	9	9(	67
%	%	%	%	%	%	%	%	%	%	%	%	%
3.1	2.5	2.8	9.9	3.1	0.3	3.4	0.3	0.6	0	2.3	0	0
6.6	3.3	4.0	8.6	7.3	0	2.6	0	0.7	0	1.7	0	0
5.9	4.5	2.8	9.4	8.7	0	2.8	0.3	0.3	0	1.0	0	0
3.0	4.3	3.5	10.9	9.8	0	2.7	0	0	0	1.9	0	0

5)	6	6(	62	63	63(	7	7(	8	8(	9	9(	67
%	%	%	%	%	%	%	%	%	%	%	%	%
2.5	7.5	6.4	4.4	4.5	0.2	1.1	0.4	0.4	0.2	1.5	0	0
4.3	10.2	4.3	3.8	4.8	0	3.8	0	0	0	1.6	1.6	0
1.8	5.9	7.9	4.9	4.7	4.3	0	1.4	0.4	0.4	0.1	1.8	0
4.9	8.9	3.5	8.2	5.5	0	0.9	0.2	0	0	0.9	0.4	0

"Memorabilia" and "Laches" than between other two dialogues of Plato.

The value of the present work is its showing of various possibilities in stylometrics, I would contend. We will, then, have to continue more experiments in order to investigate other points as a check and attempt repeating to make programs suitable for them. Only such a study can produce a general scholarly consensus about stylisitic problems.

Finally, I would like to express my gratitude to the people of the University of Manchester, especially in the Department of Greek and Latin, as well as to those in Oxford University and in the Univeersity of Edinburgh.

I wish also to acknowledge my great indebtedness to Mr. G. Neal, who kindly allowed me to bring back his elaborate programs to Japan so that I could continue my research.

Above all, I have to thank the British Council, which gave me a chance to study in Btitain.

NOTES

1) Lutoslawski, W. (1897) *The Origin and Growth of Plato's Logic.*
2) Morton, A.Q. (1971) *It's Greek to the Comuter*, p. 9.
3) Michaelson, S. (1976) *The Computation and Literary Studies*, p. 81.
4) Hockey, S. (1980) *A Guide to Computer Applications in the Humanities*, p. 41.
5) Morton, A.Q. ibid., p. 50.
6) Morton, A.Q. ibid., p. 39.
7) Kenny, A. (1976) *The Aristotelian Ethics*, p. 70.
8) Hamilton-Smith, N. (1978) *Internal Report of University of Edinburgh Department of Computer Science*, p. 2.
9) Nagase, M. (1980) "The Study of Plato by Stylometrics" *Essays and Studies*, (1981) p. 115.
10) Sentence length and position of certain words are also considered to be

effective.

11) There are some further notes relating to the coding scheme.
In case of crasis, the syntax code consists of the two separate codes (if two are required), with/between them.
+and ↑may be doubled, trebled, etc., if necessary to avoid ambiguity where there is more than one coordinate or linked pair with similar codes within a clause.
For parentheses, the appropriate *main* clause code is used after the opening bracket, and 6 is used as the Syntax Code before the bracket to indicate the relationship of the parenthetic clause to its governing clause.
The difinite article is not coded in its own right, but is given the syntax code appropriate to the word that governs it.

12) Examples are as follows:

① 101 *Πάντων μὲν οὖν μάλιστα, ὦ Σώκρατες.* ("Philebus", 11 c 4)
666 > > 66> 5> >

② 101 *Καὶ ἡμεῖς σοι τούτων γε αὐτῶν συμμάρτυρες ἂν εἶμεν, ὡς ταῦτα*
> 0> 3> 9> > 99> 4> > 1> 9(10> 2>
ἔλεγες ἃ λέγεις. ("Philebus", 12 b 3)
1> 7(15 2> 1>))

13) Morton, A.Q. ibid., pp. 55—59.

14) By checking the verbal endings like *τε*, *τεον* and *τεαν* it was reported that the usage of verbal nowns, for instance, *ὅτι φιλοσοφούντες* increased in B.C. 4th century.